Econometrica, Vol. 64, No. 5 (September, 1996), 1045–1065

MONITORING STRUCTURAL CHANGE

BY CHIA-SHANG JAMES CHU, MAXWELL STINCHCOMBE, AND HALBERT WHITE[1]

Contemporary tests for structural change deal with detections of the "one-shot" type: given an historical data set of fixed size, these tests are designed to detect a structural break within the data set. Due to the law of the iterated logarithm, one-shot tests cannot be applied to monitor out-of-sample stability each time new data arrive without signalling a nonexistent break with probability one. We propose and analyze two real-time monitoring procedures with controlled size asymptotically: the fluctuation and CUSUM monitoring procedures. We extend an invariance principle in the sequential testing literature to obtain our results. Simulation results show that the proposed monitoring procedures indeed have controlled asymptotic size. Detection timing depends on the magnitude of parameter change, the signal to noise ratio, and the location of the out-of-sample break point.

KEYWORDS: Structural change, sequential testing, fluctuation monitoring.

1. INTRODUCTION

STRUCTURAL STABILITY IS OF CENTRAL IMPORTANCE to statistical modeling of time series. In particular, if the data generating process changes in ways not anticipated by one's model, then forecasts lose accuracy. Because of the importance of structural stability, much recent effort has been devoted to obtaining convenient and powerful tests for it in a variety of modeling contexts; see, e.g., Andrews (1993), Hawkins (1987), and Ploberger, Kramer, and Kontrus (1989). The section "Breakpoints and Unit Roots" in *Journal of Business & Economic Statistics* (1992) contains work on the stability problem for nonstationary regression. However, all of this work deals with "one shot" tests: given a historical dataset of fixed size, the tests attempt to detect a structural break within the dataset.

In the real world, new data arrive steadily. Given a previously estimated model, the arrival of new data invites the question: is yesterday's model capable of explaining today's data? Breaks can occur at any point, and given the costs of failing to detect them, it is desirable to detect them as rapidly as possible. One-shot tests cannot be applied in the usual way each time new data arrive, because repeated application of such tests yields a procedure that rejects a true null hypothesis of no change with probability approaching one as the number of applications grows (Robbins (1970)). Instead, we propose a genuine sequential testing approach, yielding a procedure of controlled asymptotic size as the test is repeated.

[1] We are grateful to a co-editor and three anonymous referees for helpful comments. Any remaining errors are our responsibility. White's participation was supported by NSF Grant SES-9209023.

Wald's (1948) sequential probability ratio test (SPRT) was seminal for sequential testing. Noneconomic applications are now abundant, e.g. Page (1955) in quality control and Armitage (1975) in clinical trials. See Siegmund (1985) for an extensive bibliography. Basseville and Benveniste (1986) contains papers discussing sequential procedures from an engineering perspective. Here we use sequential testing to develop tests of structural stability for real-time monitoring of economic systems modeled using linear regression. We consider a fluctuation (FL) monitoring procedure based on recursive estimates of parameters and a CUSUM monitoring procedure based on the behavior of recursive residuals. To obtain our results, we extend certain results in the sequential testing literature.

Our sequential procedure is closely related to the theory of sequential tests with power one (Robbins (1970)). An α level sequential test of power one is a stopping rule τ such that $P(\tau < \infty \mid H_0) \leq \alpha$ and $P(\tau < \infty \mid H_1) = 1$. In one sense, our sequential procedure is a simplification of Wald's sequential procedure in view of our assumption throughout that sampling costs nothing under the null hypothesis of structural stability. In other words, we are satisfied to record the new data point without taking any action if the observed process is "in control." By contrast, Wald's sequential procedure is designed with the idea that sampling is costly under the null, as well as under the alternative. Hence, it is desirable to terminate the sequential procedure as soon as possible when the null hypothesis is true. While the assumption of costless sampling under the null is appropriate for our problem, it is not necessarily so in other applications. In fact, it is this assumption that separates us from some post-Wald sequential tests such as Anderson's (1960) sequential test of nonparallel stopping boundaries, and the truncated SPRT and repeated significance tests pioneered by Siegmund (1977). All of these tests imply $P(\tau < \infty \mid H_0) = 1$, often by construction, while this need not be true for our procedures.

This paper is organized as follows. In Section 2, we motivate and discuss the sequential testing approach. Section 3 discusses invariance principles of the past and present, and the CUSUM and fluctuation instability detectors. Section 4 contains some illustrative Monte Carlo experiments. A summary and concluding remarks are given in Section 5. Proofs are gathered into the Mathematical Appendix.

2. BACKGROUND AND MOTIVATION

2.1 *Monitoring via Repeated Retrospective Tests*

Consider a statistical model that has been estimated from a historical dataset of size m. Starting from time $m + 1$, we begin to observe fresh data sequentially, and we wish to monitor the stability of the historically adequate model. A seemingly attractive proposal is, for example, that we wait for (say) 5 periods, then perform a post-sample F test. If the F test is passed, we update the model by including these 5 new data points and wait for another 5 periods and repeat the post-sample F test; otherwise, we signal a rejection of stability. Simulation results show that thirty periods later,[2] we have a one-third chance of mistakenly signaling instability. The probability of type one error increases to 70% one hundred periods later. If the data are collected daily, ten months later we will wrongly reject the true hypothesis of stability more than 95% of the time.

Nor can retrospective tests for stability be repeatedly applied to monitor stability. To see this, we consider a simple model that will motivate all our subsequent results. For simplicity let $\{Y_t\}$ be an independent sequence. The null hypothesis of "stability in mean" is H_0: $E(Y_t) = 0$, $t = 1, 2, \ldots$, and the alternative is H_1: H_0 is false. Also assume for now that $\text{var}(Y_t) = \sigma_0^2$ for all t under H_0 and H_1.

A retrospective fluctuation test of Ploberger et al. (1989) for H_0 is given by $FL_n \equiv \max_{k \leq n} \sigma_0^{-1} \sqrt{n}\,(k/n)|\bar{Y}_k|$, where $\bar{Y}_k \equiv k^{-1} \sum_{t=1}^{k} Y_t$. The critical value c of the FL_n test is determined from the hitting probability of Brownian motion. Sequentially implementing this test leads to the stopping rule $\tau = \inf\{n \geq 1, |S_n| \geq c_n\}$, where $S_n = \sum_{t=1}^{n} Y_t$ and $c_n = \sqrt{n}\,\sigma_0 c$, i.e. when S_n exits the region bounded by $\pm c_n$, we signal instability; otherwise continue monitoring. Unfortunately, as Robbins (1970) notes, the law of iterated logarithm (LIL, e.g. Stout (1974, p. 269)) implies that under H_0, $P_0\{S_n \in [-c_n, c_n]$, for every $n \geq 1\} = 0$. Thus, sequentially implementing such a retrospective fluctuation test eventually rejects H_0, so the probability of type 1 error is one, asymptotically, and may be large in relatively small samples.

This example underscores the need to find boundary functions such that the path of S_n crosses with prescribed probability under H_0. Such a choice is possible using the functional central limit theorem (FCLT) to approximate the boundary crossing probabilities of a normalized partial sum by those of standard Brownian motion.

A limiting relation that is central to our approach is

$$(1) \qquad \lim_{m \nearrow \infty} P\left\{S_n \geq \sqrt{m}\, g(n/m), \text{ for some } n \geq 1\right\}$$
$$= P\{W(t) \geq g(t), \text{ for some } t \geq 0\},$$

where $S_n = \sum_{t=1}^{n} \varepsilon_t$, W denotes a standard Brownian motion, and g is a stopping boundary satisfying some regularity conditions. This relation was first proved by Robbins and Siegmund (1970) for iid $\{\varepsilon_t\}$. Our monitoring procedures are based on an extension of Robbins and Siegmund's theorem that holds generally for sums S_n obeying a FCLT, under milder restrictions on g. The use of Brownian motion in sequential testing has two merits. First, it is convenient, offers some qualitative insight, and provides a unified treatment for many problems. Second, it handles the technical difficulty arising from the overshoot problem in a discrete time framework since the excess over the stopping boundary in continuous time tends to be stochastically small.

[2] In the simulation, we generated data from $y_t = 0.6x_t + \varepsilon_t$ where ε_t is niid(0, 1) and x_t is an AR(1) with the AR coefficient 0.8. The in-sample size is 270; nominal size is 5%. The number of replications is 4,000.

2.2 *Stability and Regression*

Our interest throughout focuses on linear regression $Y_t = X_t' \beta_t + \varepsilon_t$, $t = 1, 2, \ldots$, where X_t is a $k \times 1$ random vector and β_t is a $k \times 1$ nonstochastic vector. Throughout, we make the following "noncontamination" assumption:

ASSUMPTION A: $\beta_t = \beta_0$ *for* $t = 1, 2, \ldots, m$.

Thus, the regression parameters are stable over the historical period of length m. We are interested in the hypothesis of regression stability in the "post-historical" period, H_0: $\beta_t = \beta_0$, $t = m+1, \ldots$, versus the alternative H_1 of instability that β_t changes at some $t \geq m+1$.

A monitoring scheme is a stopping time, determined by a detecting statistic (detector) Γ_n and a threshold $g(m,n)$, according to $\tau_g(\Gamma_n) \equiv \min\{n \geq m, \Gamma_n > g(m,n)\}$. The traditional choice of the detector is the likelihood ratio statistic (LR); however we propose two other detectors, the cumulative sum (CUSUM) of recursive residuals and the parameter fluctuations (FL). The relation between the CUSUM recursive residuals and other CUSUM algorithms in the engineering literature as well as the relation between LR and FL detector will be discussed below.

Consider first the CUSUM detector. Let $\hat{\beta}_n = (\sum_{i=1}^n X_i X_i')^{-1} (\sum_{i=1}^n X_i Y_i)$ be the OLS estimator at time n. Define recursive residuals as $\omega_k = 0$ and $\omega_n = \hat{\varepsilon}_n / v_n^{1/2}$, $v_n = 1 + X_n' (\sum_{i=1}^{n-1} X_i X_i')^{-1} X_n$, $\hat{\varepsilon}_n = Y_n - X_n' \hat{\beta}_{n-1}$, $n = k+1, \ldots m, \ldots$. The nth cumulated sum of recursive residuals is $Q_t^m = \hat{\sigma}^{-1} \sum_{i=k}^n \omega_i = \hat{\sigma}^{-1} \sum_{i=k}^{k+[\tilde{m}t]} \omega_i$, for $(n-k)/\tilde{m} \leq t < (n-k+1)/\tilde{m}$, where $\hat{\sigma}$ is a consistent estimator of σ, $\tilde{m} = (m-k)$, and $[\tilde{m}t]$ is the integer part of $\tilde{m}t$. It is well known (e.g., Kramer, Ploberger, and Alt (1988)) that under H_0

$$\{t \to \tilde{m}^{-1/2} Q_t^m, t \in [0, \infty)\} \Rightarrow \{t \to W(t), t \in (0, \infty)\}, \tag{2}$$

where "$\Rightarrow$" denotes the weak convergence of the associated probability measures.

As the monitoring starts at $m+1$, define $\tilde{Q}_t^m = \hat{\sigma}^{-1} \sum_{i=m+1}^{k+[\tilde{m}(1+t)]} \omega_i$, $t \in [0, \infty)$. In particular, for $n/(m-k) \leq t < (n+1)/(m-k)$, $\tilde{Q}_n^m = \hat{\sigma}^{-1} (\sum_{i=k}^{m+n} \omega_i - \sum_{i=k}^m \omega_i)$, $n \geq 1$. It follows that $t \to \tilde{m}^{-1/2}\, \tilde{Q}_t^m \Rightarrow t \to [W(t+1) - W(1)]$, $t \in [0, \infty)$. The limiting process $W(t+1) - W(1)$ is a Brownian motion. If (1) holds, then we have (given (5) and (6) below)

$$\lim_{m \nearrow \infty} P\left\{ |\tilde{Q}_n^m| \geq \sqrt{m-k}\, g\left(\frac{n}{m-k} \right), \text{for some } n \geq 1 \right\}$$
$$= P\{|W(t)| \geq g(t), \text{for some } t \geq 0\}.$$

This suggests the following monitoring procedure: whenever the path of $|\tilde{Q}_n^m|$ exits the boundary $(m-k)^{1/2} g(n/m-k)$, the null hypothesis is rejected, implying that the model identified from the stable historical period is no longer relevant for forecasting.

Another limiting result similar to (1) is that

$$\lim_{m \nearrow \infty} P\left\{|S_n| \geq \sqrt{m}\, g(n/m), \text{ for some } n \geq m\right\} = P\{|W(t)| \geq g(t), \text{ for some } t \geq 1\}. \tag{3}$$

This result is central to the later FL monitoring procedure; it can be applied to (2) to yield an alternative CUSUM monitoring scheme as

$$\lim_{m \nearrow \infty} P\left\{|Q_n^m| \geq \sqrt{m-k}\, g\left(\frac{n}{m-k}\right), \text{ for some } n \geq m\right\} = P\{|W(t)| \geq g(t), \text{ for some } t \geq 1\}. \tag{4}$$

CUSUM algorithms other than those using recursive residuals can also be considered. For a simple null versus simple alternative hypothesis, the CUSUM algorithm based on LR statistics for independent sequences is well known, e.g. Page (1955) and Hinkley (1971). The Page-Hinkley test is optimal in the sense of minimizing the conditional detection delay (Lorden (1971), Moustakides (1986)). However, except for very simple models, the LR detector is exceedingly complex and its optimality is no longer available in general (in Section 3.4 we take up optimality issues in greater detail). From an engineering perspective, some simplifications of the computationally heavy LR detector are desirable. Nikiforov (1986) proposes using an asymptotic expansion of LR detector to reduce the complexity. The resulting CUSUM of scores is shown to be effective against local alternatives. Basseville (1986) studies on-line change detection in autoregressive models and suggests alternative detectors based on Kullback divergence and Chernoff's distance between conditional laws. Our CUSUM is not a LR-type detector in general, but its computational advantage is obvious, which makes it easy to build into software. In contrast, Basseville's approach has to be studied separately for each problem and is difficult to study theoretically. As for the choice of the threshold, Nikiforov's CUSUM algorithm uses a constant threshold which needs to be fine tuned to ensure proper test size and short detection delay (Nikiforov (1986, pp. 245–248)); we use the invariance principle to suggest a particular stopping boundary, perhaps a less sophisticated but nevertheless quite convenient approach.

An alternative approach to monitoring stability rests upon monitoring the stability of the fluctuations of sequential parameter estimates. We take this up in Section 3.3 below, where we show that convenient tests for stability can again be constructed using relation (3). The foregoing discussion thus motivates us to seek conditions that will permit application of (1) or (3) whenever S_n satisfies a FCLT, under conditions on g as mild as possible. We now turn to this task.

3. MONITORING PARAMETER CHANGE

3.1 *Invariance Principles and Boundary Crossing Probabilities*

Let $\{S_n\}$ be the partial sum process constructed from $\{\varepsilon_i\}$. We say that $\{\varepsilon_i\}$ follows the FCLT if $\lambda \to \sigma_0^{-1} m^{-1/2} S_{[m\lambda]} \Rightarrow \lambda \to W(\lambda)$, $\lambda \in [0, \infty)$, where $S_n =$

$\Sigma_{i=1}^n \varepsilon_i$ and $\sigma_0^2 \equiv \lim_{n \nearrow \infty} n^{-1}E(S_n^2) < \infty$. Robbins and Siegmund (1970) proved that for an iid $\{\varepsilon_i\}$ (3) holds for a certain class of continuous functions $g(t)$ that satisfy (a) $t^{-1/2}g(t)$ is ultimately nondecreasing as $t \to \infty$, and (b) $\int_\tau^\infty t^{-3/2}g(t)\exp[-(1/2t)g^2(t)]dt < \infty$. In addition to (a) and (b), (1) also holds under (c) $t^{-1/2}g(t)$ is nonincreasing for t sufficiently small, and (d) $\int_0^1 t^{-3/2}g(t)\exp[-(1/2t)g^2(t)]dt < \infty$.

Though this is useful, the iid assumption is not always applicable in time series regression, where the errors may be martingale differences and not independent. To make our monitoring scheme more relevant for time series applications, it is useful to have the same limit relation hold for a wider class of stochastic sequences than just iid. Segen and Sanderson (1980) provide a partial resolution of this problem. Under the condition that the boundary function $t^{-1/2}g(t)$ is nondecreasing, they show that for any stochastic sequence $\{S_n\}$ that satisfies the FCLT the limit relation (1) continues to hold. However, this condition on g is more restrictive than the condition that $t^{-1/2}g(t)$ is ultimately nondecreasing imposed by Robbins and Siegmund. As a result, convenient boundary functions such as the ones we use in Section 3.3 below are ruled out. Further, without the nondecreasing assumption for $t^{-1/2}g(t)$, Segen and Sanderson's proof will not work.

We now propose a limiting relation similar to Segen and Sanderson's without imposing the nondecreasing condition. For this, let D be the union of the set of continuous functions on $[0,\infty)$ and the set of cadlag piecewise constant functions on $[0,\infty)$ such that for all $f \in D$, f has at most finitely many discontinuities over any interval $[0, N]$, $N \in \mathbb{R}^+$, and such that every discontinuity point of f is rational. Endow D with the metric

$$d(f,g) = \sum_{n \in \mathbb{N}} 2^{-n} \min\left\{ \sup_{t \in [0,n]} |f(t) - g(t)|, 1 \right\}.$$

This induces the topology of uniform convergence on compacta on D. Let $C \subseteq D$ denote the set of continuous functions on $[0,\infty)$. Let μ be the Wiener measure on $(D, \mathscr{D})$, where $\mathscr{D}$ is the Borel σ-field on D. Note that $\mu(C) = 1$.

Let (Ω, F, P) be a complete probability space and $W: \Omega \times [0,\infty) \to \mathbb{R}$ be measurable on Ω and continuous on $[0,\infty)$. Suppose that the mapping $\omega \to W(\omega, .)$ induces Wiener measure on $(D, \mathscr{F})$. Given $g: [0,\infty) \to \overline{\mathbb{R}}^+$ (the extended real numbers), define the epigraph epi $g = \{(t,x): x \geq g(t)\}$ and epi $g^+ = \{(t,x): x > g(t)\}$. For a measurable, real-valued stochastic process X on $\Omega \times [0,\infty)$, define $\tau_g^x = \inf\{t \geq 0: X(.,t) \in \text{epi } g\}$, and $\tau_{g^+}^x = \inf\{t \geq 0: X(.,t) \in \text{epi } g^+\}$. We say that $\tau_g^{\cdot}$ is continuous at $f \in D$ if $d(f^n, f) \to 0$ implies that $\rho(\tau_g^f, \tau_g^{f^n}) \to 0$, where ρ is a metric on $[0,\infty]$ inducing the Euclidean topology on $[0,\infty]$ and having the property that any sequence increasing without bound converges to ∞. We use the class of "regular" functions g having two properties:

(5) $\quad P\left\{\tau_g^W = \tau_{d^+}^W\right\} = 1$ and

(6) $\quad P\{\omega: \tau_t^{\cdot}$ is continuous at $W(\omega, .)\} = 1$.

These two properties for g are fairly abstract. The following lemma provides some useful sufficient conditions; see Appendix A for a proof.

LEMMA 3.1: *The class of regular g includes the class of cadlag functions from* $[0,\infty)$ *to* $\overline{\mathbb{R}}^+$ *having* $g(0)>0$, *and having at most countably many points*, t, *such that the upper right derivative of g at t is infinity.*

Let $\tau_{1,g}^X = \inf\{t \geq 1: X(\omega,t) \in \text{epi } g\}$, and $\tau_{1,g^+}^X = \inf\{t \geq 1: X(\omega,t) \in \text{epi } g^+\}$. Small modifications in the proof of Lemma 3.1 allow us to dispense with the condition $g(0)>0$. We record this as Lemma 3.2 without proof.

LEMMA 3.2: *The class of g such that* $P\{\tau_{1,g}^X = \tau_{1,g^+}^X\} = 1$ *and* $P\{\tau_{1,g}^{\cdot}$ *is continuous at* $W(\omega,.)\} = 1$ *includes the class of cadlag functions from* $[0,\infty)$ *for* $\overline{\mathbb{R}}^+$ *having at most countably many points t such that the upper right derivative of g at t is infinity.*

Let $\{S_n\}_{n\in\mathbb{N}}$ be a sequence of random variables defined on $(\Omega,\mathscr{F},P)$ with $S_0=0$. For each $\omega\in\Omega$, define $\{X_t^0(\omega): t\in[0,\infty)\}$ by $X_t^0(\omega)=S_{[t]}(\omega)$. For each $m\in\mathbb{N}$ define $X^m=\{X_t^m: t\in[0,\infty)\}$ by $X_t^m = m^{-1/2}X_{mt}^0$. Let μ^m be the probability measure on $(D,\mathscr{D})$ induced by X^m (that is, $\mu^m(A)=P\{\omega: X^m(\omega)\in A\}$).

THEOREM 3.3: *Suppose that* $\mu^m \Rightarrow \mu$. *If g is a regular function, then* $\tau_g^{x^m}$ *converges to* τ_g^W *in distribution, and* $\tau_{1,g}^{X^m}$ *converges in distribution to* $\tau_{1,g}^W$.

Because the $[0,\infty]$-valued random variables $\tau_g^{X^m}$ are converging to τ_g^W in distribution only, we cannot conclude (without further argument) that $P\{\tau_g^{X^m}<\infty\}\to P\{\tau_g^W<\infty\}$. However, we can conclude for all T in a set with an at most countable complement in $[0,\infty)$, $P\{\tau_g^{X^m}<T\}\to P\{\tau_g^W<T\}$ and $P\{\tau_{1,g}^{X^m}<T\}\to P\{\tau_{1,g}^W<T\}$. This is surely sufficient for sensible applications, but for those concerned about extreme tail behavior, we include an extra condition on g's behavior near infinity that guarantees $P\{\tau_g^{X^m}<\infty\}\to P\{\tau_g^W<\infty\}$ and $P\{\tau_{1,g}^{X^m}<\infty\}\to P\{\tau_{1,g}^W<\infty\}$. The condition ensures that $P\{\tau_g^W=\infty\}=0$.

THEOREM 3.4: *Let* $\{S_n,\ n\in\mathbb{N}\}$ *be a stochastic sequence defined on a complete probability space. Suppose that* (*i*) *the probability measure* μ^m *on the measurable space of cadlag functions* $(D,\mathscr{D})$ *induced by* $X_t^m \equiv \{m^{-1/2}S_{[mt]},\ t\in[0,\infty)\}$ *converges weakly to the Wiener measure* μ, *i.e.* $m^{-1/2}S_{[mt]}\Rightarrow W(t)$, $t\in[0,\infty)$, (*ii*) *g is a regular function, and* (*iii*) $t^{-1/2}g(t)$ *is eventually nondecreasing. Then*:

(*a*) $\lim_{m\nearrow\infty} P\{S_n\geq\sqrt{m}\,g(n/m)$, *for some* $n\geq 1\} = P\{W(t)\geq g(t)$, *for some* $t\geq 0\}$;

(*b*) $\lim_{m\nearrow\infty} P\{S_n\geq\sqrt{m}\,g(n/m)$, *for some* $n\geq m\} = P\{W(t)\geq g(t)$, *for some* $t\geq 1\}$.

Thus (1) holds when S_n satisfies a FCLT and g is mildly restricted. Theorem 3.4 supports application to testing regression stability in a range of time-series contexts relevant to economics.

We close this section with remarks on the calculation of the boundary crossing probability. The computation is generally difficult in the sense that analytical expressions for the boundary crossing probability are not always available for an arbitrary function $g(t)$. Robbins and Siegmund (1970) give a useful theorem (see Theorem A in the Appendix) powerful enough to cover many well known results. Instructive examples are given in an earlier version of this paper (Chu, Stinchcombe, and White (1993)). Sen (1981) also gives some miscellaneous results. An extensive treatment is given by Lerche (1984). Two particular boundary functions considered in the sequel are

$$P\left\{|W(t)| \geq \left[t(a^2 + \ln t)\right]^{1/2}, \text{for some } t \geq 1\right\} \qquad (7)$$
$$= 2[1 - \Phi(a) + a\phi(a)] \quad \text{and}$$

$$P\left\{|W(t)| \geq (t+1)^{1/2}\left[a^2 + \ln(t+1)\right]^{1/2}, \text{for some } t > 0\right\} \qquad (8)$$
$$= \exp(-a^2/2),$$

where Φ and ϕ are the cdf and pdf respectively of a standard normal random variable.

3.2 *Application to CUSUM Monitoring*

The results of the previous section now can be immediately applied to obtain a CUSUM monitoring result. It suffices to pick g properly. Unfortunately, this choice is often dictated by mathematical convenience rather than optimality, since crossing probabilities for an arbitrary boundary are analytically intractable in general. Segen and Sanderson (1980) choose $g(t) = 2t\log(\log t)$, a function satisfying the nondecreasing condition in their theorem. Due to the LIL, this $g(t)$ increases about as slowly as is possible to have $P\{|W(t)| \geq g(t)$, for some $t \geq 0\} < 1$. Hence, Segen and Sanderson's choice is motivated by the fastest detection of change.

We shall not consider this boundary since our simulations suggest that its probability of type one error is erratic in finite samples. Moreover, the choice of $g(t) = 2t\log(\log t)$ highlights the main difference between much of (but not all of) the engineering and quality control applications and our economic situation. Elsewhere it is implicitly or explicitly assumed that the system can be reset at small or zero cost after a false alarm. We practically never have that option in economics. Hence, we use (7) and (8) to implement a CUSUM monitoring, summarized as follows.

COROLLARY 3.5: *Suppose (i) $Y_t = X_t'\beta_0 + \varepsilon_t$, $t = 1, \ldots, m+1, \ldots$, where X_t is a $k \times 1$ random vector such that $m^{-1}\sum_{t=1}^m X_t$ and $m^{-1}\sum_{t=1}^m X_t X_t'$ converge in probability to b, a nonstochastic $k \times 1$ vector and M, a $k \times k$ matrix of full rank, respectively; (ii) $\{\varepsilon_t\}$ is a martingale difference sequence with respect to a sequence of σ-algebras $\{F_t\}$ such that $E(\varepsilon_t^2) < \infty$ and $E(\varepsilon_t^2 \mid F_{t-1}) = \sigma_0^2$ for all t, where F_t is generated by $\{\ldots, (Y_{t-2}, X_{t-1}'), (Y_{t-1}, X_t')\}$; (iii) the sequence $\{X_t \varepsilon_t\}$ obeys the* *functional central limit theorem, i.e.* $\lambda \to m^{-1/2}(\sigma_0^2 M)^{-1/2}\sum_{t=1}^{[m\lambda]} X_t \varepsilon_t \Rightarrow \lambda \to W(\lambda)$, $\lambda \in [0,\infty)$. *Then*:

$$\lim_{m \nearrow \infty} P\left\{|\tilde{Q}_n^m| \geq \sqrt{n+m-k}\left[a^2 + \ln\left(\frac{n+m-k}{m-k}\right)\right]^{1/2}, \text{ for some } n \geq 1\right\} \tag{9}$$

$$= \exp(-a^2/2);$$

$$\lim_{m \nearrow \infty} P\left\{|Q_n^m| \geq \sqrt{m-k}\left(\frac{n}{m-k}\right)^{1/2}\left[a^2 + \ln\left(\frac{n}{m-k}\right)\right]^{1/2}, \text{ for some } n \geq m\right\} = 2[1 - \Phi(a) + a\phi(a)]. \tag{10}$$

The required conditions in Corollary 3.5 are not the weakest, but they are sufficient to cover many interesting cases. The asymptotic sizes of the CUSUM monitoring are fairly easy to control from the right-hand side of (9) and (10). For CUSUM monitoring based on $\tilde{Q}_n^m$, the 10% and 5% asymptotic size corresponds to $a^2 = 4.6$ and 6 respectively. Equally handily, we obtain the 10% and 5% asymptotical size of the CUSUM monitoring based on Q_n^m by setting $a^2 = 6.25$ and 7.78 in (10) respectively.

Consider the alternative of a one-time parameter shift $H_1: Y_t = X_t'\beta_0 + \varepsilon_t$, $t = 1, 2, \ldots [m\tau]$, and $Y_t = X_t'\beta_1 + \varepsilon_t$, $t = [m\tau] + 1, \ldots,$ where $\tau > 1$ is the break point. The CUSUM monitoring procedure can be shown to be consistent if the mean regressor is not orthogonal to the magnitude of shift $(\beta_1 - \beta_0)$. Since $E(\hat{\varepsilon}_i) = 0$ for $i \leq [m\tau]$ and $E(\hat{\varepsilon}_i) \cong (\tau/\lambda)E(X_i')(\beta_1 - \beta_0)$ for $i > [m\tau]$ and $\lambda > \tau$, consistency can be established by showing that

$$\lim_{m \nearrow \infty} ([\tilde{m}\lambda] + k)^{-1}\left|\sigma_0^{-1}\sum_{i=k}^{[\tilde{m}\lambda]+k} \omega_i\right| < \infty, \quad \text{while}$$

$$\lim_{m \nearrow \infty} ([\tilde{m}\lambda] + k)^{-1/2}\{a^2 + \ln(([\tilde{m}\lambda] + k)/\tilde{m})\}^{1/2} = 0.$$

Before ending this section, we emphasize that it is practically possible to implement the CUSUM monitoring with other nontrivial boundaries. Since a boundary function is defined implicitly through the measure F in Theorem A, one first defines the measure F in a way that the results of Theorem A hold. Of course, we should not expect to obtain neat analytical solutions as in (9) or (10), but we can resort to such programs as MATHEMATICA to evaluate the integral in the right side of Theorem A(b) numerically by specializing h, b, and τ.

3.3 *Fluctuation Monitoring*

An alternative approach rests upon monitoring the stability of the fluctuations of sequential parameter estimates. Let $Y_t = X_t'\beta_0 + \varepsilon_t$, as before. The key condition is that the sequence $\{X_t \varepsilon_t\}$ obeys the multivariate functional central limit theorem (Phillips and Durlauf (1986), Wooldridge and White (1988)):

$$\lambda \to m^{-1/2} V_0^{-1/2} \sum_{t=1}^{[m\lambda]} X_t \varepsilon_t \Rightarrow \lambda \to \underline{W}(\lambda), \quad \lambda \in [0, \infty),$$

where $V_0 = \lim_{m \to \infty} m^{-1} E(S_m S_m')$ with $S_m = \Sigma_{t=1}^m X_t \varepsilon_t$, and $\underline{W}(\lambda)$ is a k-dimensional Wiener process such that each elements of $\underline{W}(\lambda)$ is a univariate Wiener process, independent of the others.

We define a fluctuation detector by

$$(11) \qquad \hat{Z}_n = n D_m^{-1/2} \left(\hat{\beta}_n - \hat{\beta}_m \right), \qquad n \geq m,$$

where $D_m = M_m^{-1} V_0 M_m^{-1}$, M_m is $O(1)$ and uniformly positive definite such that $(\Sigma_{t=1}^m X_t X_t'/m) - M_m \xrightarrow{p} 0$. The essential ingredient of this FL detector is the deviation of the updated parameter estimate $\hat{\beta}_n$ from the historical parameter estimate $\hat{\beta}_m$. If the null hypothesis is correct, the process of constant interpolation of $\hat{Z}_n$ is in control and all of the component process of $\hat{z}_n^i$, $i = 1, \ldots, k$, will stay below a monitoring boundary $g(n/m)$ with probability $1 - \alpha$, i.e. $P_0\{|\hat{z}_n^i| < m^{1/2} g(n/m),\ n = m+1, \ldots,$ for all $i = 1, \ldots, k\} = 1 - \alpha$.

Define $\Psi: D^k \to D^k$ as a continuous functional such that $\Psi(f) = f(t) - tf(1)$. Let

$$(12) \qquad Z_n^0 = n D_m^{-1/2} \left(\hat{\beta}_n - \beta_0 \right),$$

and $X_\lambda^m = m^{-1/2} Z_{[m\lambda]}^0$, so that

$$X_\lambda^m = m^{-1/2} D_m^{-1/2} \left(\sum_{t=1}^{[m\lambda]} X_t X_t' / [m\lambda] \right)^{-1} \sum_{t=1}^{[m\lambda]} X_t \varepsilon_t .$$

Directly form the multivariate FCLT, $X_\lambda^m \Rightarrow \underline{W}(\lambda)$. Hence, $\Psi(X_\lambda^m) = m^{-1/2} \hat{Z}_{[m\lambda]} \Rightarrow \Psi(\underline{W}(\lambda)) = \underline{W}(\lambda) - \lambda \underline{W}(1) \equiv \underline{W}^0(\lambda)$, $\lambda \in [1, \infty)$. Since the limiting process, $\{\underline{W}^0(\lambda),\ \lambda \in [1, \infty)\}$, is a k-dimensional Brownian bridge, with elements stochastically independent of one another, $\lim_{m \nearrow \infty} P_0\{|\hat{z}_n^i| < m^{1/2} g(n/m),\ n = m+1, \ldots,$ for all $i = 1, \ldots, k\} = [P\{|W^0(\lambda)| < g(\lambda),\ \lambda \geq 1\}]^k$. From this it suffices to consider the boundary crossing probabilities of a univariate Brownian bridge process.

The process $\{W^0(\lambda),\ \lambda \in [1, \infty)\}$ has covariance $E[W^0(t) W^0(s)] = t(s-1)$, for $t > s > 1$. It can be verified that $\{W^0(\lambda),\ \lambda \in [1, \infty)\} \stackrel{d}{=} \{(\lambda - 1) W(\lambda/\lambda - 1),\ \lambda \in [1, \infty)\}$ by rescaling the time parameter. It follows that $P\{W^0(\lambda) \geq g(\lambda)$, for some $\lambda \geq 1\} = P\{(\lambda - 1) W(\lambda/\lambda - 1) \geq g(\lambda)$, for some $\lambda \geq 1\}$. Hence the crossing probability of $W^0(\lambda)$ can be investigated in terms of the Wiener process. We choose $g(t) = [t(a^2 + \ln t)]^{1/2}$ as in (7), which is analytically convenient but not necessarily optimal in the sense of minimal detection delay. Putting $t = \lambda/\lambda - 1$, it follows that

$$(13)\quad P\left\{(t-1)^{-1}|W(t)| \geq (t-1)^{-1}[t(a^2+\ln t)]^{1/2}, \text{ for some } t \geq 1\right\}$$
$$= P\left\{(\lambda-1)\left|W\left(\frac{\lambda}{\lambda-1}\right)\right| \geq (\lambda-1)\left[\left(\frac{\lambda}{\lambda-1}\right)\left[a^2+\ln\left(\frac{\lambda}{\lambda-1}\right)\right]\right]^{1/2}, \text{ for some } \lambda \geq 1\right\}$$
$$= 2[1-\Phi(a)+a\phi(a)].$$

The crossing probability can be easily computed via the right side of (13). When $a^2 = 7.78$ and 6.25, the crossing probabilities are 0.05 and 0.10 respectively. Discretizing λ as n/m yields the boundary

$$g(n/m) = \left(\frac{n-m}{m}\right)\left[\left(\frac{n}{n-m}\right)\left[a^2+\ln\left(\frac{n}{n-m}\right)\right]\right]^{1/2}.$$

It follows that

$$\lim_{m\nearrow\infty} P\left\{|\hat{z}^i_{[m\lambda]}| \geq m^{1/2}\left(\frac{n-m}{m}\right)\left[\left(\frac{n}{n-m}\right)\left[a^2+\ln\left(\frac{n}{n-m}\right)\right]\right]^{1/2}, \text{ for some } n \geq m\right\} \cong 2[1-\Phi(a)+a\phi(a)].$$

The foregoing discussions can be summarized in the following corollary.

COROLLARY 3.6: *Suppose that (i) $Y_t = X_t'\beta_0 + \varepsilon_t$; (ii) $\{X_t\varepsilon_t\}$ obeys a FCLT with $V_0 = \lim_{m\nearrow\infty} m^{-1}E[(\Sigma_{t=1}^m X_t\varepsilon_t)(\Sigma_{t=1}^m X_t\varepsilon_t)']$ positive definite; (iii) $m^{-1}\Sigma_{t=1}^m X_tX_t' - M_m \xrightarrow{p} 0$, where $\{M_m\}$ is $O(1)$ and uniformly positive definite; (iv) there exists a positive semi-definite matrix $\hat{D}_m$ such that $\hat{D}_m - D_m \xrightarrow{p} 0$, where $D_m = M_m^{-1}V_0M_m^{-1}$.*

Let $\hat{Z}_n = nD_m^{-1/2}(\hat{\beta}_n - \hat{\beta}_m)$. Then:

(a) $\lambda \to m^{-1/2}\hat{Z}_{[m\lambda]} \Rightarrow \lambda \to \underline{W}^0(\lambda)$, $\lambda \in [1,\infty)$, where $\underline{W}^0(\lambda)$ is a k-dimensional Brownian bridge:

$$(b)\quad \lim_{m\nearrow\infty} P\left\{|\hat{z}^i_{[m\lambda]}| \geq m^{1/2}\left(\frac{n-m}{m}\right)\left[\left(\frac{n}{n-m}\right)\left[a^2+\ln\left(\frac{n}{n-m}\right)\right]\right]^{1/2}, \text{ for some } n \geq m \text{ and some } i\right\}$$
$$= 1-[1-2[1-\Phi(a)+a\phi(a)]]^k,$$

where $\hat{z}^i_{[m\lambda]}$ is the ith component of $\hat{Z}_{[m\lambda]}$.

Corollary 3.6 summarizes the monitoring procedure based on the fluctuation of $\hat{\beta}_n$ $(n > m)$ relative to $\hat{\beta}_m$. Given the number of regressors, k, and arbitrary probability of type one error α, we can determine the constant a^2 in the monitoring boundary. Note that $\hat{\beta}_m$ is never updated. This, however, does not mean that we always use $\hat{\beta}_m$ for subsequent analysis. If there is no warning of parameter shift, it is $\hat{\beta}_n$ that we would use for forecasting rather than $\hat{\beta}_m$, since $\hat{\beta}_n$ is a more accurate estimate than $\hat{\beta}_m$ under the null hypothesis.

The FL monitoring in Corollary 3.6 has asymptotic power one under global alternatives of one-time parameter shift; we can show that

$$m^{-1/2}\left(\frac{n-m}{m}\right)\left[\left(\frac{n}{n-m}\right)\left[a^2+\ln\left(\frac{n}{n-m}\right)\right]\right]^{1/2} \to 0$$

while $\max_i\{m^{-i}|z^i_{[m\lambda]}|\}$ diverges. Formally we have the following theorem:

THEOREM 3.7: *Suppose* (*i*) $Y_t = X_t'\beta_0 + \varepsilon_t$ *for* $t = 1,2,\ldots[m\tau]$, *and* $Y_t = X_t'\beta_1 + \varepsilon_t$ *for* $t = [m\tau]+1,\ldots$, $\beta_1 \neq \beta_0$, $\tau > 1$; (*ii*) *all conditions in Corollary* 3.6 *hold except* (*a*). *Then*

$$\lim_{m\nearrow\infty} P\left\{|\hat{z}^i_{[m\lambda]}| \geq m^{1/2}\left(\frac{n-m}{m}\right)\left[\left(\frac{n}{n-m}\right)\left[a^2+\ln\left(\frac{n}{n-m}\right)\right]\right]^{1/2},\right.$$
$$\left.\text{for some } n \geq m \text{ and some } i\right\} = 1.$$

3.4 *Discussion*

Because most prior results focus on the univariate location model, we specialize our results accordingly to make a direct comparison with the sequential testing literature. Let X_t be unity. The random function in (12) becomes $Z^0_n = n\sigma_0^{-1}(\bar{Y}_n - \beta_0)$. It follows that

$$\text{(14)} \quad \lim_{m\to\infty} P\left\{|Z^0_n| \geq (n+m)^{1/2}[a^2+\log(n/m+1)]^{1/2}, \text{ for some } n \geq 1\right\}$$
$$= P\left\{|W(t)| \geq (t+1)^{1/2}[a^2+\ln(t+1)]^{1/2}, \text{ for some } t > 0\right\}$$
$$= \exp(-a^2/2).$$

Because β_0 is known here, m does not necessarily correspond to the historical sample size. It is perhaps useful to think of m as determining the resolution of the constant interpolation constructed from the $\{Z^0_n\}$ sequence. To start monitoring from $n = 1$, we track the path of the FL detector Z^0_n every time a new data point arrives. Note that the continuous version of this monitoring boundary is $g(t) = (t+1)^{1/2}[a^2 + \ln(t+1)]^{1/2}$, which does not satisfy the requirement that $t^{-1/2}g(t)$ is nondecreasing over $[0,\infty)$. Hence (14) is not a consequence of Segen and Sanderson's (1980) theorem.

The accuracy of the asymptotic approximation of error probability in (14) clearly depends on m. Following Robbins (1970) by setting $m = 1$ and $a^2 = 4.6$, we obtain $g(n/m) = g(n) = (n+1)^{1/2}[4.6 + \ln(n+1)]^{1/2}$ giving size less than 10% (see the further derivation below). One can also choose larger m, say $m = 100$ to obtain $g(n/100) = (n+100)^{1/2}[4.6 + \ln((n/100)+1)]^{1/2}$, which should deliver better finite sample size than the choice $m = 1$. However, the diameter of the acceptable region for the monitoring boundary with $m = 1$ is smaller (greater) than that with $m = 100$, provided $n \leq 133$ ($n > 133$). Consequently, if the parameter shift occurs earlier (later), the choice $m = 1$ provides faster (slower) detection than the choice of $m = 100$, though both will detect it eventually. We have seen that the performance of FL monitoring in terms of the detection delay depends on the choice of m, which is in turn determined by the location of the break point. Hence, the optimal m cannot be determined without an explicit alternative about the location of a future break point.

Another interesting aspect of (14) is its relation to the traditional Wald SPRT. Consider a sequential test for H_0: $\beta_0 = 0$ against the normal mixture alternative H_1: $\beta_0 = \delta$, such that $dF(\rho) = (2\pi)^{-1/2}\exp(-\rho^2/2)d\rho$, where $\rho = \delta/\sigma_0$ is the magnitude of change relative to the standard deviation of the noise. Let h_n^0 and h_n^1 be the pdf's of $(Y_1, \ldots, Y_n)$ under H_0 and H_1 respectively. The LR is $l_n = h_n^2/h_n^0$, where $h_n^2 = \int_{-\infty}^{\infty} h_n^1\, dF(\rho)$. Straightforward algebra yields

$$\begin{aligned} l_n &= \int \exp\left(-\tfrac{1}{2}n\rho^2 + \rho Z_n^0\right) dF(\rho) \\ &= (2\pi)^{-1/2} \int \exp\left[-\tfrac{1}{2}(n+1)\rho^2 + \rho Z_n^0\right] d\rho, \text{where } Z_n^0 = n\sigma_0^{-1}\bar{Y}_n. \end{aligned}$$

But

$$\left[-\tfrac{1}{2}(n+1)\rho^2 + \rho Z_n^0\right] = -\tfrac{1}{2}(n+1)\left[\rho - Z_n^0/n+1\right]^2 + \frac{1}{2(n+1)} Z_n^{0^2},$$

so we have

$$l_n = (n+1)^{-1/2}\exp\left[\frac{1}{2(n+1)} Z_n^{0^2}\right].$$

It follows from Wald's lemma (1948, p. 146) that $P\{l_n \geq \varepsilon, \text{for some } n \geq 1\} \leq \varepsilon^{-1}$. Specializing to $\varepsilon = \exp(a^2/2)$ gives $P\{|Z_n^0| \geq (n+1)^{1/2}[a^2 + \log(n+1)]^{1/2}$, for some $n \geq 1\} \leq \exp(-a^2/2)$. This is precisely the monitoring boundary in (14) when $m = 1$. Of course, this kind of interpretation is not necessarily available for other boundary functions. In fact, Wald's SPRT under different alternatives will result in completely different monitoring boundaries.

Consider now H_0: $\beta = \beta_0$ vs. H_1: $\beta = \beta_1$. It is easy to show that Wald's SPRT with the stopping rule $\tau^* = \inf\{n \geq 1, l_n \geq c\}$ yields a linear monitoring boundary $c/\sigma_0 + [(\beta_1 - \beta_0)/2\sigma_0]n$. It is well known that Wald's SPRT in the simple vs. simple hypothesis is optimal in the sense that for all stopping rules such that $P(\tau | H_0) \leq \alpha$, τ^* minimizes the expected delay $E(\tau | H_1)$: see Siegmund (1985). Hence the present linear boundary is optimal while the monitoring boundary in

(14) is suboptimal. Nevertheless, this remarkable optimal property applies only to the case of simple hypotheses and an independent error sequence. If the alternative hypothesis is composite, one should expect to obtain a different boundary. For example, in the previous example of the normal mixture alternative, we followed Wald's SPRT principle and used the same stopping rule $\inf\{n \geq 1,\ l_n \geq c\}$, but obtained a different boundary. To assert some sort of optimal property for (14) (say, in the sense of minimal delay) is a challenging task. In fact, there is no known optimality result in the composite hypothesis testing. One can only hope that Wald's SPRT principle with its associated stopping rule continues to perform well in composite hypothesis testing situations.

When β_0 is unknown, the relevant random function becomes $\hat{Z}_n = n\sigma_0^{-1}(\bar{Y}_n - \bar{Y}_m)$; see (11). In contrast to the well developed literature of the usual two-sample sequential test, we are now dealing with the so-called partially sequential two-sample test, perhaps a less well known branch of the sequential testing literature; see Wolfe (1977) and Switzer (1983). In the partially sequential two-sample framework, the first sample is fixed while the second sample is observed sequentially after the first sample is complete. In our framework, the first sample is the historical sample of fixed size m, a data set available to econometricians at the time of designing the FL monitoring. The second sample is observed sequentially, and we do not wish to take more observations than are necessary to signal instability.[3] In one sense, our FL monitoring mimics Wald's SPRT since it reduces to the Wald's SPRT when $\bar{Y}_m$ is replaced with β_0. In another sense, our FL monitoring is a modification of Switzer's procedure. To see this, rewrite $\hat{Z}_n = \sigma_0^{-1}(n-m)(\bar{Y}_{n-m} - \bar{Y}_m)$, where $\bar{Y}_{n-m} \equiv 1/(n-m)\sum_{t=m+1}^{n} Y_t$. Switzer's (1983) stopping rule is: $\tau = \inf\{n\colon |\bar{Y}_{n-m} - \bar{Y}_m| \geq c/n-m\}$, which is just the FL detector with constant boundary, i.e. $\tau = \inf\{n\colon |\hat{Z}_n| \geq c\}$, assuming $\sigma_0 = 1$. Clearly, Switzer's procedure terminates with probability one even when H_0 is true. This is not appropriate since we assume that sampling costs nothing under H_0.

To investigate the optimality of the monitoring boundary, it seems attractive to treat m as fixed but make the choice of m part of the monitoring design scheme. Intuitively, the choice of m affects the average run length (ARL, i.e. $E(\tau \mid H_1)$) in the second sample. Switzer (1983) considered the optimal choice of m for independent sequences. While an analytical solution is not available, his simulations seem to indicate the existence of optimal choice. We do not have an optimal choice for m at present; further research on this topic is warranted.

[3] If β_0 is known, the partially sequential two-sample test reduces to the one-sample sequential test such as Wald's SPRT. If one does not care about the detection timing and can afford to wait until enough post-m observations are in, post-historical stability can be investigated via nonsequential retrospective stability tests, which is the case of the fixed two-sample test. If m is treated as random rather than given (such as in a controlled experiment), we have the standard two-sample sequential test.

4. SIMULATIONS

In this section, we simulate the finite sample properties of our monitoring procedures. To focus our discussion, we concentrate on FL monitoring; results on CUSUM monitoring are not reported here but are available on request. We generate data from iid $N(2,1)$ random variables and compute the empirical crossing probabilities under H_0 for historical sample sizes $m = 25$, 50, 100, 200, and 300. The monitoring horizon q is set to be two, four, six, and nine times the historical sample size. Theory predicts that if m is large enough and q is extended to infinity, the crossing probabilities should approach 5% and 10% if a^2 in Corollary 3.6 is 7.78 and 6.25 respectively. Though it is not possible to set q equal to infinity in simulations, the results summarized in Table I give no hint of improper size except when $m = 25$. The empirical crossing probabilities when $m = 25$ are slightly over the 5% nominal level, indicating that the probabilities when $q = \infty$ will be overstated.

To examine the power of FL monitoring, we create an artificial out-of-sample structural break at $t = 1.1 \times m$, at which the mean shifts from 2 to 2.8 permanently. The FL monitoring indeed signals the structural change eventually, i.e. the test is consistent. Traditionally, one is more concerned about the ARL of a sequential procedure. For this, we summarize the empirical distribution of the first hitting time in Table II. Three remarks are in order.

First, the standard deviation of the first hitting time decreases significantly from $m = 25$ to 50 and from $m = 50$ to 100. Moreover, the standard deviation is rather stable after m reaches 100. This seems to suggest that the precision of FL monitoring depends upon how accurately we can estimate the unknown parameter from the historical sample. Intuitively, the parameter estimate from the historical sample serves as the benchmark in the FL monitoring; consequently, increased accuracy of parameter estimates resulting from larger m improves the monitoring precision. The fact that the standard deviation stabilizes after $m = 100$ is perhaps peculiar to our simulated location model. A sample of 100 should provide a reliable estimate of the mean; the accuracy gain from a sample of size larger than 100 is insignificant in this scenario. One conjecture is that increasing the historical sample size beyond a certain point does not appreciably improve the precision of detection.

TABLE I

EMPIRICAL SIZES OF FL MONITORING

	$m = 25$		$m = 50$		$m = 100$		$m = 200$		$m = 300$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$q = 2m$	4.0	8.0	3.5	6.7	3.0	6.0	3.3	7.0	3.2	6.8
$= 4m$	5.1	8.6	4.1	7.0	3.7	7.1	3.4	7.4	3.6	7.4
$= 6m$	5.2	9.5	4.3	7.6	4.2	7.6	3.6	7.5	3.5	7.7
$= 9m$	5.2	9.6	4.3	8.0	4.3	7.9	4.0	8.0	4.1	7.9

Note: The number of replications is 2,500.

TABLE II

EMPIRICAL FIRST HITTING TIME

	Break	1st Q	Median	3rd Q	Mean	Std	ARL	MRL
$m = 25$	28	37	46	65	64	58	36	490
$= 50$	55	66	74	88	82	30	27	434
$= 100$	110	124	131	142	135	16	25	184
$= 125$	138	153	161	172	164	16	26	139
$= 150$	165	181	189	199	191	15	26	144
$= 200$	220	237	246	256	248	15	28	110
$= 250$	275	294	302	312	305	15	28	115
$= 300$	330	350	360	370	361	16	31	104

Note: The number of replications is 2,500, with a 10% monitoring boundary. The first quantile (1st Q) is the integer that gives empirical probability closest to 25% rather than exactly 25%; similarly for the median and third quantile. Mean, standard deviation (std), ARL, and MRL are rounded to nearest integers.

Second, Table II also shows the ARL and maximal detection delay (MRL) of FL monitoring. The ARL's are more or less stable except when m is small ($m = 25$). The MRL signifies the worst performance in terms of detection timing. The relation between MRL and m resembles a continuous decreasing function. The benefit of having a shorter MRL in large samples seems significant. It is reasonable that these measures of detection timing (ARL and MRL) depend on the magnitude of parameter change. Simulation results under the alternative of larger magnitude of change (not reported here) indeed confirms this conjecture. Specifically, both the ARL and MRL decrease as expected. When $m = 100$ and the parameter shifts from 2 to 3, the ARL decreases to 17 (it was 25 in Table II), and the MRL decreases from 184 to 85. Moreover, the standard deviation of first hitting time is also found to be lower as the magnitude of parameter change gets larger.

Third, it is interesting to note that the distribution of the first hitting time is generally nonsymmetric for smaller samples, indicated by the discrepancy of the mean and the median. Nevertheless, the asymmetry becomes less obvious when the historical sample size increases.

We also simulate and record the empirical first hitting time distributions for an out-of-sample structural break occurring at $t = 1.2m$. The FL monitoring procedure is found to be less effective. All of the ARL, MRL, and standard deviations increase. In particular, given the same magnitude of change, the ARL is prolonged to 42 periods when $m = 100$. Ideally, we would like to have the same ARL regardless of the location of the break point. This ideal situation does not occur because the monitoring boundary we chose increases a little too fast. Here we have encountered an obvious dilemma. The limiting process for fluctuation monitoring, $W^0(\lambda)$, has growing variance; hence the monitoring boundary must be increasing to take into account this fact. How slow a growth rate in the boundary function can be allowed in order to have the right test size and faster detection is dictated by the computation of the boundary crossing

TABLE III
ESTIMATION OF THE BREAK POINT

History(m)	$m = 100$		$m = 200$		$m = 300$	
True break	110	120	220	240	330	360
Avg estimate	114	122	222	238	330	356
Std error	8	8	7	9	7	11

probabilities. The growing variance of $W^0(\lambda)$, which induces an increasing monitoring boundary, is a problem inherent to this type of monitoring. Even if the crossing probability for an increasing monitoring boundary with slower growth rate than the one we currently use can be computed, the ARL will still be longer in the alternative of a late structural break. One way to improve the detection timing is to consider a fixed window monitoring rather than the growing window (recursive) FL monitoring. This approach is considered by Chu, Hornik, and Kuan (1995), in which a retrospective moving-estimates test is proposed. Another possibility is to adjust m as we move through time. We leave treatment of this possibility to later work.

After the FL monitoring procedure signals a structural change, the next step is to revise the model. To do so, it is necessary to know the location of the break point. We suggest using the point at which the maximum of the LR statistics is obtained, defined from time point $m+1$ to the first hitting time (see Horvath (1994)). The performance of this locating procedure is briefly summarized in Table III. It is seen that the maximum LR statistics do pretty well in locating the break point.

5. CONCLUDING REMARKS

Any statistical model, no matter how well it fits the historical data, must always face the challenge: is yesterday's model capable of explaining today's data? An historically adequate model that behaves poorly outside the data set cannot generate accurate forecasts. We have suggested two real-time monitoring procedures for high frequency data sets: the CUSUM and the FL monitoring procedures. We also widen the class of boundary functions beyond those suggested in the literature. Our simulation results conform well with the intuition that detection timing depends on the magnitude of parameter change, the standard deviation of the disturbance term, and the location of the break point.

We have focused on deriving monitoring procedures with asymptotically correct size for a given boundary function. It is obvious that the choice of the boundary function determines the speed of detection. We chose a particular monitoring boundary for mathematical convenience. Of all the continuous boundary functions that have everywhere finite upper right derivative, there may exist a choice that gives an optimal monitoring in the sense of minimal delay. Addressing this difficult issue is left for further research.

Dept. of Economics, MC 0253, University of Southern California, Los Angeles, CA 90089, U.S.A.,

Dept. of Economics, University of Texas at Austin, Austin, TX 78712, U.S.A.,

and

Dept. of Economics, D-008, University of California—San Diego, La Jolla, CA 92093, U.S.A.

Manuscript received October, 1993; final revision received December, 1994.

APPENDIX

THEOREM A (Robbins and Siegmund (1970)): *Let $W(.)$ be a standard Wiener process on $[0,\infty)$ and F be any measure, finite on bounded intervals. Define $f(x,t) \equiv \int_0^\infty \exp(\theta x - \theta^2 t/2)F(d\theta)$, $x, t \in R$, and $A(t,\varepsilon) \equiv \inf\{x: f(x,t) \geq \varepsilon\}$, $\varepsilon \in (0,\infty)$. It follows that $x < A(t,\varepsilon)$ implies $f(x,t) < \varepsilon$, and $f(x,t) < \varepsilon$ implies $x \leq A(t,\varepsilon)$. Furthermore, if for some $b, h, f(b,h) < \infty$, then for each $t > h$, $f(x,t) = \varepsilon$ has unique solution $x = A(t,\varepsilon)$. Let ϕ and Φ be the pdf and cdf of the standard normal.*

(a) *For any b, h, ε such that $f(b,h) < \varepsilon$, $P\{W(t) > A(t+h,\varepsilon) - b$, for some $t > 0\} = \varepsilon^{-1} f(b,h)$.*

(b) *For any b, h, ε and $\tau > 0$,*

$$\begin{aligned} &P\{W(t) > A(t+h,\varepsilon) - b, \text{for some } t > \tau\} \\ &\quad = 1 - \Phi[\tau^{-1/2}[A(\tau+h,\varepsilon) - b]] \\ &\qquad + \varepsilon^{-1}\int_0^\infty \exp(\theta b - \theta^2 h/2)\Phi[\tau^{-1/2}[A(\tau+h,\varepsilon) - b] - \theta\tau^{1/2}]\,dF(\theta). \end{aligned}$$

If the measure $F(.)$ defined over $\mathbb{R}$ is symmetric, the conclusion remains valid if $W(t)$ is replaced by $|W(t)|$. This theorem is general in the sense that for various g the crossing probabilities can be computed by choosing an appropriate measure F. Note that in (1) $g(t)$ is just $A(t+h,\varepsilon)$ with given h, ε, and τ is set to one.

PROOF OF LEMMA 3.1: Let g denote a function satisfying the given assumptions. Define a Borel measure Q on $[0,\infty]$ by $Q(A) = P\{\tau_g^W \in A\}$. Note that $Q\{\infty\} > 0$ is possible.

Step A: If $t > 0$ is a continuity point of g, then $Q\{t\} = 0$.

Because $N(0,t)$ is a smooth distribution, it suffices to show that $\{\tau_g^W = t\} \subset \{W(.,t) = g(t)\}$. Pick $\omega \in \{W(.,t) \neq g(t)\}$; we will show that $\omega \notin \{\tau_g^W = t\}$. First suppose that $W(\omega,t) - g(t) = \varepsilon > 0$. Because g and $W(\omega,.)$ are continuous at t, there exist a $\delta_1 > 0$ and a $\delta_2 > 0$ such that $|t - t'| < \delta_1$ implies $|g(t) - g(t')| < \varepsilon/3$ and that $|t - t'| < \delta_2$ implies $|W(\omega,t) - W(\omega,t')| < \varepsilon/3$. Let $\delta = \min\{\delta_1, \delta_2\}$. For all $t^* \in (t-\delta, t)$, $W(\omega,t^*) > g(t^*)$. Hence $\tau_g^W(\omega) < t$ so that $\omega \notin \{\tau_g^W = t\}$. When $g(t) - W(\omega,t) = \varepsilon > 0$, a similar proof shows that $\tau_g^W(\omega) \neq t$.

Step B: If $t > 0$ is a point of discontinuity of g and g jumps upwards at t, that is $g(t) > \lim_{s \nearrow t} g(s)$, then $Q(t) = 0$.

Because g is right continuous, $\{\tau_g^W = t\} \subset \{W(.,t) = g(t)\}$ and $P\{W(.,t) = g(t)\} = 0$.

Step C: If $t > 0$ is a point of discontinuity of g and $a := g(t) < b := \lim_{s \nearrow t} g(s)$, and $Q\{t\} = 0$, then $P\{W(.,t) \in (a,b) \mid \tau_g^W = t\} = 1$.

By the smoothness of normal distributions, it suffices to show that if $\omega \in \{\tau_g^W = t\}$, then $W(\omega,\tau) \leq b$ and $W(\omega,t) \geq a$. If $W(\omega,t) > b$, then the definition of b and the continuity of $W(\omega,.)$

imply that there exists a δ such that for all $t'' \in (t-\delta, t)$, $W(\omega, t'') > g(t'')$. Hence $\tau_g^W(\omega) < t$. Similarly, if $W(\omega, t) < a$, then $\tau_g^W(\omega) \neq t$.

We now prove (5). Since $\operatorname{epi} g^+ \subset \operatorname{epi} g$ implies that $\tau_g^W \leq \tau_{g^+}^W$ for all ω, if $\tau_g^W(\omega) = \infty$, then $\tau_{g^+}^W(\omega) = \infty$. The only remaining case is $\tau_g^W < \infty$. We divide $\{\tau_t^W < \infty\}$ into two sets, depending on whether or not $\tau_g^W(\omega)$ belongs to the set of discontinuities of g.

Now the discontinuities of g, denoted Δ, form a countable, hence measurable set. By Step A, Q must be nonatomic on the complement of the discontinuities of g. By assumption, the points of g with infinite upper right derivative are countable, hence their intersection with Δ^c has Q-mass zero. By the strong Markov property, for every $N \in \mathbb{N}$, the process $B^N(\omega, t) = W(\omega, t + (\tau_g^W(\omega) \wedge N)) - W(\omega, \tau_g^W(\omega) \wedge N)$ is a Brownian motion. It is well known that with probability one, the upper right derivative of $B^N(\omega, .)$ at $t = 0$ is infinity. Combining these observations, we have that outside of a set of measure zero, if $\tau_g^W(\omega) = \tau_g^W(\omega) \wedge N$ and $W(\omega, .)$ first hits g at a point in Δ^c, then it hits where $W(\omega, .)$ has an infinite upper right derivative. For such ω, $\tau_g^W(\omega)$ must equal $\tau_{g^+}^W(\omega)$. Letting N tend to infinity shows that for almost all ω in $\{\tau_g^W(\omega) < \infty\} \cap \{\tau_t^W(\omega) \in \Delta^c\}$, $\tau_g^W(\omega) = \tau_{g^+}^W(\omega)$.

Because Δ is measurable, Step B implies that for almost all ω in $\{\tau_g^W \in D\}$ and for all $t \in \Delta$, $\tau_g^W(\omega) = t$ implies $g(t) < \lim_{s \nearrow t} g(s)$. Step C and continuity of $W(\omega, .)$ imply that, outside a set of measure zero, every ω in $\{\tau_g^W \in \Delta\}$ satisfies $\tau_g^W(\omega) = \tau_{g^+}^W(\omega)$. Thus, excepting at most a set of measure zero, we have $\tau_g^W = \tau_{g^+}^W$, proving (5).

We now turn to (6). The crucial step is the following:

Step D: For all $r > 0$, and for almost all ω in $\{\tau_g^W > r\}$, there exists a $\delta > 0$ such that $\inf\{|W(\omega, t) - g(t)|: t \in [0, r]\} > \delta$.

Let $T_n(\omega)$ denote the set $\{t \in [0, r]: |\bar{g}(t) - W(\omega, t)| \leq 1/n\}$, where $\bar{g}$ is the closed graph correspondence defined by $\bar{g}(t) = \{g(t), g(t-)\}$, where $g(t-) = \lim_{s \nearrow t} g(s)$. $T_n(\omega)$ is a closed subset of the compact set $[0, r]$, and $T_{n+1}(\omega) \subset T_n(\omega)$. Hence either there exists an n' such that for all $n \geq n'$, $T_n(\omega) = \varnothing$, or $T(\omega) = \cap_n T_n(\omega)$ is nonempty. Suppose $T(\omega)$ is not empty, and let $t^0 \in T(\omega)$. If t^0 is a continuity point of g, that is, $t^0 \in \Delta^c$, then $W(\omega, t^0) = g(t^0)$, which implies that $\tau_g^W(\omega) \leq r$, a contradiction. If $t^0 \in \Delta$, then the continuity of $W(\omega, .)$ implies that $W(\omega, t^0) = g(t^0)$ or $g(t^0 -)$. But because the set Δ is countable, $P(\bigcup_{t \in \Delta}\{\omega: W(\omega, t) \in \{g(t), g(t-)\}\}) = 0$.

We now prove the almost everywhere continuity of $\tau_g^{\cdot}$. By (5), we need only prove the almost everywhere continuity of $\tau_{g^+}^{\cdot}$. Because $g(0) > 0$ and g is right continuous, the LIL implies that $P\{\tau_{g^+}^W = 0\} = 0$. We will divide the continuity proof into two parts, $\tau_{g^+}^W \in (0, \infty)$ and $\tau_{g^+}^W = \infty$.

Suppose that $\tau_{g^+}^W(\omega) = r \in (0, \infty)$ and X^n is a sequence of functions with $d(X^n, W(\omega, .)) \to 0$. We must show that for all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n \geq N$, $|\tau_{g^+}^W(\omega) - \tau_{g^+}^{X^n}| < \varepsilon$. Pick arbitrary $\varepsilon > 0$. By step D, there exists a $\delta_1 > 0$ such that $\inf\{|W(\omega, t) - g(t)|: t \in [0, r - \varepsilon/2]\} > 2\delta_1$. By the definition of $\tau_{g^+}^W$, there exists a $\delta_2 > 0$ and a $t'' \in [r, r + \varepsilon/2]$ such that $W(\omega, t'') > g(t'') + 2\delta_2$. Let $\delta = \min\{\delta_1, \delta_2\}$. Pick N sufficiently large that for all $n \geq N$,

$$\sup\{|X^n(t) - W(\omega, t)|: t \in [0, r - \varepsilon/2]\} < \delta.$$

By the definition of δ_1, $\tau_{g^+}^{X^n} \geq r - \varepsilon/2 > r - \varepsilon$. By the definition of δ_2, $\tau_{g^+}^{X^n} \leq r + \varepsilon/2 < r - \varepsilon$. Hence $|\tau_{g^+}^W - \tau_{g^+}^{X^n}| < \varepsilon$ as required.

Now suppose that $\tau_{g^+}^W(\omega) = \infty$, and X^n is a sequence of functions with $d(X^n, W(\omega, .)) \to 0$. We must show that for all $r \in (0, \infty)$, there exists an $N \in \mathbb{N}$ such that for all $n \geq N$, $\tau_{g^+}^{X^n} > r$. By Step D, there exists a $\delta > 0$ such that $\inf\{|W(\omega, t) - g(t)|: t \in [0, r+1]\} > \delta$. Pick N large so that for all $n \geq N$, $\sup\{|X^n(t) - W(\omega, t)|: t \in [0, r+1]\} < \delta$. For all $n \geq N$, $\tau_{g^+}^W > r$. This completes the entire proof of Lemma 3.1. *Q.E.D.*

Proof of Theorem 3.3: The results follow by the continuous mapping theorem. *Q.E.D.*

PROOF OF COROLLARY 3.6: Let $\tilde{Z}_{[m\lambda]} = [m\lambda]D_m^{-1/2}(\hat{\beta}_{[m\lambda]} - \hat{\beta}_m)$ and $\tilde{z}_{i,[m\lambda]}$ be the ith element of $\tilde{Z}_{[m\lambda]}$. It follows that

$$P\Big\{|\tilde{z}_{i,n}| < m^{1/2}(n-m/m)[(n/n-m)[a^2+\ln(n/n-m)]]^{1/2}, \text{ for all } n \geq m \text{ and all } i\Big\}$$
$$= 1 - \Big[1 - P\Big\{|\tilde{z}_{i,n}| \geq m^{1/2}(n-m/m)[(n/n-m)[a^2+\ln(n/n-m)]]^{1/2},$$
$$\text{for some } n \geq m\Big\}\Big]^k.$$

But

$$\lim_{m\nearrow\infty} P\Big\{|\tilde{z}_{i,n}| \geq m^{1/2}(n-m/m)[(n/n-m)[a^2+\ln(n/n-m)]]^{1/2}$$
$$= P\{(\lambda-1)|W(\lambda/\lambda-1)| \geq (\lambda-1)[(\lambda/\lambda-1)[a^2+\ln(\lambda/\lambda-1)]]^{1/2},$$
$$\text{for some } \lambda \geq 1\} = 2[1-\Phi(a)+a\phi(a)],$$

from (13). The above conclusion remains valid if we replace D_m with its consistent estimator. *Q.E.D.*

PROOF OF THEOREM 3.7: Let $\lambda > \tau > 1$.

$$\hat{\beta}_{[m\lambda]} = \left(\sum_{t=1}^{[m\lambda]} X_t X_t'\right)^{-1}\left(\sum_{t=1}^{[m\lambda]} X_t Y_t\right) = \left(\sum_{t=1}^{[m\lambda]} X_t X_t'\right)^{-1}\left(\sum_{t=1}^{[m\tau]} X_t Y_t + \sum_{t=[m\tau]+1}^{[m\lambda]} X_t Y_t\right)$$
$$= \left(\sum_{t=1}^{[m\lambda]} X_t X_t'\right)^{-1}\left(\sum_{t=1}^{[m\tau]} X_t X_t'\right)\hat{\beta}_{1,[m\tau]} + \left(\sum_{t=1}^{[m\lambda]} X_t X_t'\right)^{-1}$$
$$\times\left(\sum_{t=[m\tau]+1}^{[m\lambda]} X_t X_t'\right)\hat{\beta}_{[m\tau]+1,[m\lambda]},$$

where $\hat{\beta}_{1,[m\tau]}$ is the estimate from observation 1 to $[m\tau]$ and $\hat{\beta}_{[m\tau]+1,[m\lambda]}$ from observation $[m\tau]+1$ to $[m\lambda]$. It follows that $(\hat{\beta}_{[m\lambda]} - \hat{\beta}_m) \to (1-(\tau/\lambda))\delta$, where $\delta = (\beta_1 - \beta_0)$, and that

$$\lim_{m\to\infty} \sup_{\lambda} \frac{1}{m}\hat{Z}_{[m\lambda]} = \lim_{m\to\infty} \sup_{\lambda} \frac{[m\lambda]}{m} D_m^{-1/2}\left(\hat{\beta}_{[m\lambda]} - \hat{\beta}_m\right)$$

diverges. On the other hand, $m^{-1/2}(n-m/m)[(n/n-m)[a^2+\ln(n/n-m)]]^{1/2} \to 0$, and the theorem follows. *Q.E.D.*

REFERENCES

ANDERSON, T. W. (1960): "A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size," *Annals of Mathematical Statistics*, 31, 165–197.

ANDREWS, D. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ARMITAGE, P. (1975): *Sequential Medical Trials*, 2nd ed. Oxford: Blackwell.

BASSEVILLE, M. (1986): "The Two-Models Approach for the On-line Detection of Changes in Autogressive Process," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-Verlag Lecture Notes in Control and Information Sciences 77, ed. by M. Basseville and A. Benveniste. New York: Springer-Verlag, pp. 169–215.

BASSEVILLE, M., AND A. BENVENISTE (1986): *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-Verlag Lecture Notes in Control and Information Sciences 77. New York: Springer-Verlag.

CHU, C. S. J., K. HORNIK, AND C. M. KUAN (1995): "A Moving-Estimates Test for Parameter Stability," *Econometric Theory*, 11, 699–720.

CHU, C. S. J., M. STINCHCOMBE, AND H. WHITE (1993): "Monitoring Structural Change," UCSD, Department of Economics Discussion Paper.

HAWKINS, D. (1987): "A Test for a Change Point in a Parametric Model Based on a Maximal Wald-type Statistic," *Sankhya*, 49, 368–376.

HINKLEY, D. (1971): "Inference about the Change-Point from Cumulative Sum Tests," *Biometrika*, 58, 509–523.

HORVATH, L. (1994): "Detecting Changes in Linear Regressions," University of Utah, Department of Mathematics Working Paper.

Journal of Business and Economic Statistics (1992): Special Section—"Break Points and Unit Roots," 237–335.

KRAMER, W., W. PLOBERGER, AND R. ALT (1988): "Testing for Structural Change in Dynamic Models," *Econometrica*, 56, 1355–1370.

LERCHE, H. (1984): *Boundary Crossing Probabilities of Brownian Motion*. New York: Springer-Verlag.

LORDEN, G. (1971): "Procedures for Reacting to a Change in Distribution," *Annals of Mathematical Statistics*, 42, 1897–1908.

MOUSTAKIDES, G. (1986): "Optimal Stopping Times for Detecting Changes in Distributions," *Annals of Statistics*, 14, 1379–1387.

NIKIFOROV, I. (1986): "Sequential Detection of Changes in Stochastic Systems," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-Verlag Lectures Notes in Control and Information Sciences 77, ed. by M. Basseville and A. Benveniste. New York: Springer-Verlag, pp. 216–258.

PAGE, E. (1955): "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42, 523–527.

PHILLIPS, P., AND S. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

PLOBERGER, W., W. KRAMER, AND K. KONTRUS (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

ROBBINS, H. (1970): "Statistical Methods Related to The Law of the Iterated Logarithm," *Annals of Mathematical Statistics*, 41, 1397–1409.

ROBBINS, H., AND D. SIEGMUND (1970): "Boundary Crossing Probabilities for the Wiener Process and Sample Sums," *The Annals of Mathematical Statistics*, 41, 1410–1429.

SEGEN, J., AND A. SANDERSON (1980): "Detection of Changes in a Time Series," *IEEE Transaction in Information Theory*, IT-26, 249–257.

SEN, P. K. (1981): *Sequential Nonparametrics: Invariance Principle and Statistical Inference*. New York: Wiley & Sons.

SIEGMUND, D. (1977): "Repeated Significance Tests for a Normal Mean," *Biometrika*, 64, 177–189.

——— (1985): *Sequential Analysis*. New York: Springer-Verlag.

STOUT, W. F. (1974): *Almost Sure Convergence*. New York: Academic Press.

SWITZER, P. (1983): "A Two-Sample Sequential Test for Shift with One Sample Size Fixed in Advance," in *Recent Advances in Statistics*, edited by M. Haseeb Rizvi, Jagdish S. Rustagi, and David Siegmund. New York: Academic Press.

WALD, A. (1948): *Sequential Analysis*. New York: Wiley & Sons.

WOLFE, D. (1977): "On a Class of Partially Sequential Two-Sample Test Procedures," *Journal of American Statistical Association*, 72, 202–205.

WOOLDRIDGE, J., AND HALBERT WHITE (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.

LINKED CITATIONS

Monitoring Structural Change
Chia-Shang James Chu; Maxwell Stinchcombe; Halbert White
Econometrica, Vol. 64, No. 5. (Sep., 1996), pp. 1045-1065.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199609%2964%3A5%3C1045%3AMSC%3E2.0.CO%3B2-Z

References

A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size
T. W. Anderson
The Annals of Mathematical Statistics, Vol. 31, No. 1. (Mar., 1960), pp. 165-197.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28196003%2931%3A1%3C165%3AAMOTSP%3E2.0.CO%3B2-N

Tests for Parameter Instability and Structural Change With Unknown Change Point
Donald W. K. Andrews
Econometrica, Vol. 61, No. 4. (Jul., 1993), pp. 821-856.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199307%2961%3A4%3C821%3ATFPIAS%3E2.0.CO%3B2-I

Inference about the Change-Point from Cumulative Sum Tests
D. V. Hinkley
Biometrika, Vol. 58, No. 3. (Dec., 1971), pp. 509-523.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197112%2958%3A3%3C509%3AIATCFC%3E2.0.CO%3B2-T

Searching for a Break in GNP
Lawrence J. Christiano
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 237-250.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C237%3ASFABIG%3E2.0.CO%3B2-8

LINKED CITATIONS

Testing for Structural Change in Dynamic Models
Walter Krämer; Werner Ploberger; Raimund Alt
Econometrica, Vol. 56, No. 6. (Nov., 1988), pp. 1355-1369.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198811%2956%3A6%3C1355%3ATFSCID%3E2.0.CO%3B2-4

Procedures for Reacting to a Change in Distribution
G. Lorden
The Annals of Mathematical Statistics, Vol. 42, No. 6. (Dec., 1971), pp. 1897-1908.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28197112%2942%3A6%3C1897%3APFRTAC%3E2.0.CO%3B2-G

Optimal Stopping Times for Detecting Changes in Distributions
George V. Moustakides
The Annals of Statistics, Vol. 14, No. 4. (Dec., 1986), pp. 1379-1387.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198612%2914%3A4%3C1379%3AOSTFDC%3E2.0.CO%3B2-%23

A Test for a Change in a Parameter Occurring at an Unknown Point
E. S. Page
Biometrika, Vol. 42, No. 3/4. (Dec., 1955), pp. 523-527.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28195512%2942%3A3%2F4%3C523%3AATFACI%3E2.0.CO%3B2-K

Multiple Time Series Regression with Integrated Processes
P. C. B. Phillips; S. N. Durlauf
The Review of Economic Studies, Vol. 53, No. 4, Econometrics Special Issue. (Aug., 1986), pp. 473-495.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28198608%2953%3A4%3C473%3AMTSRWI%3E2.0.CO%3B2-R

Statistical Methods Related to the Law of the Iterated Logarithm
Herbert Robbins
The Annals of Mathematical Statistics, Vol. 41, No. 5. (Oct., 1970), pp. 1397-1409.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28197010%2941%3A5%3C1397%3ASMRTTL%3E2.0.CO%3B2-Y

LINKED CITATIONS

Boundary Crossing Probabilities for the Wiener Process and Sample Sums
Herbert Robbins; David Siegmund
The Annals of Mathematical Statistics, Vol. 41, No. 5. (Oct., 1970), pp. 1410-1429.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28197010%2941%3A5%3C1410%3ABCPFTW%3E2.0.CO%3B2-B

Repeated Significance Tests for a Normal Mean
D. Siegmund
Biometrika, Vol. 64, No. 2. (Aug., 1977), pp. 177-189.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197708%2964%3A2%3C177%3ARSTFAN%3E2.0.CO%3B2-T